

# Tripartitions do not always discriminate phylogenetic networks\*

**Gabriel Cardona**

Department of Mathematics  
and Computer Science  
University of the Balearic Islands  
E-07122 Palma de Mallorca  
Spain

**Francesc Rosselló**

Research Institute of Health Science  
University of the Balearic Islands  
E-07122 Palma de Mallorca  
Spain

**Gabriel Valiente**

Algorithms, Bioinformatics, Complexity  
and Formal Methods Research Group  
Technical University of Catalonia  
E-08034 Barcelona  
Spain

February 1, 2008

## Abstract

Phylogenetic networks are a generalization of phylogenetic trees that allow for the representation of non-tree-like evolutionary events, like recombination, hybridization, or lateral gene transfer. In a recent series of papers devoted to the study of reconstructibility of phylogenetic networks, Moret, Nakhleh, Warnow and collaborators introduced the so-called *tripartition metric* for phylogenetic networks. In this paper we show that, in fact, this tripartition metric does not satisfy the separation axiom of distances (zero distance means isomorphism, or, in a more relaxed version, zero distance means indistinguishability in some specific sense) in any of the subclasses of phylogenetic networks where it is claimed to do so. We also present a subclass of phylogenetic networks whose members can be singled out by means of their sets of tripartitions (or even clusters), and hence where the latter can be used to define a meaningful metric.

**Keywords.** Phylogenetic networks, recombination, bipartitions, tripartitions, tripartition metric, error metric

## 1 Introduction

Phylogenetic trees have been used since the days of Darwin [4] to represent evolutionary histories of sets of species under mutation. Their popularity and prevalence have led to the introduction of many methods to their reconstruction, combination, and comparison [6, 8, 19]. But, as Doolittle pointed out about a decade ago [7], the history of life cannot be properly represented as a tree. Phylogenetic networks are used then as a generalization of

---

\*This work has been partially supported by the Spanish CICYT project TIN2004-07925-C03-01 GRAMARS, by Spanish DGI projects MTM2006-07773 COMGRIO and MTM2006-15038-C02-01, and by EU project INTAS IT 04-77-7178.

phylogenetic trees that allow for the representation of non-treelike evolutionary events, like recombination, hybridization, or lateral gene transfer [1].

The natural model for describing an evolutionary history is a directed acyclic graph (DAG for short) representing the parent-child relation. Phylogenetic trees are rooted DAGs where each node other than the root (which represents the common ancestor of all individuals under consideration, be them species or biomolecular sequences) has at most one parent, from which it has been derived through mutation. Phylogenetic networks are rooted DAGs containing *tree nodes*, which have only one parent and thus correspond to regular speciation events, and *hybrid nodes*, which have more than one parent and thus correspond to hybrid speciation events. To such a DAG several extra conditions have been added in the literature to provide a realistic model of recombination [20, 21] or simply to narrow the output space of reconstruction algorithms [9, 10].

In a series of papers [11, 12, 13, 14, 15, 16, 17] devoted to the study of reconstructibility of phylogenetic networks, Moret, Nakhleh, Warnow and collaborators introduced a method to compare a reconstructed network and the true phylogeny, the so-called *tripartition*, or also *error*, *metric*. This method is based on the association, to each node  $v$  of the network, of a tripartition of its set of leaves into those that are strict descendants of  $v$  (that is, such that every path from the root to the leaf contains  $v$ ), those that are non-strict descendants of it (that is, that are descendants but not strict descendants), and those that are not descendants of  $v$ . These tripartitions may be enriched with some extra information like, for instance, the greatest number of hybrid nodes in a path from  $v$  to each leaf, or the sets of descendants of the parents of hybrid nodes. Notice anyway that these tripartitions are a natural generalization to non-tree networks of Bourque-Robinson-Foulds bipartitions, which associate to each node of a phylogenetic tree the partition of its leaves into descendant and non-descendant ones. These bipartitions are used to define one of the most popular distances for phylogenetic trees [3, 18].

One of the key points in the definition of this tripartition metric is the claim that the sets of tripartitions discriminate, up to isomorphism, phylogenetic networks in a suitable subclass of them. That is, that if two phylogenetic networks  $N$  and  $N'$  in this subclass have the same sets of tripartitions, then they are isomorphic. This turns out to be equivalent to the separation axiom for the tripartition metric (zero distance means isomorphism). In this paper, we provide counterexamples showing that all claims made in this connection in those papers are untrue, and thus that the tripartition metric does not satisfy the separation axiom in any of the cases considered by the authors, even in the more relaxed sense of [14], where zero distance is simply claimed to be equivalent to equality modulo a certain specific notion of indistinguishability. Therefore, the tripartition metric cannot be used in a meaningful way to compare phylogenetic networks in the classes considered in the aforementioned papers, as it cannot decide the equality (or even indistinguishability) of networks. Then, in the last section, we show a slight variant of the class considered in one of these papers where tripartitions, and even bipartitions in the sense of Bourque-Robinson-Foulds, do define a metric.

## 2 Notations on DAGs

Let  $N = (V, E)$  be a directed acyclic graph (DAG). We denote by  $d_i(u)$  and  $d_o(u)$  the in-degree and out-degree, respectively, of a node  $u \in V$ .

A node  $v \in V$  is a *leaf* if  $d_o(v) = 0$ . A node  $v \in V$  is a *tree node* if  $d_i(v) \leq 1$ . Such a tree node is a *root* if  $d_i(v) = 0$ , and *internal* if  $d_i(v) = 1$  and  $d_o(v) > 0$ . A node  $v \in V$  is *hybrid* if  $d_i(v) > 1$ . We denote by  $V_L$ ,  $V_T$ , and  $V_H$  the sets of leaves, of tree nodes, and of

hybrid nodes of  $N$ , respectively. An arc  $(u, v) \in E$  is a *tree arc* if its head  $v$  is a tree node, and a *network arc* if  $v$  is hybrid.

A *clade* of a DAG  $N$  is a subtree of  $N$  with set of nodes contained in  $V_T$  and set of leaves contained in  $V_L$ .

A node  $v \in V$  is a *child* of  $u \in V$  if  $(u, v) \in E$ ; we also say that  $u$  is a *parent* of  $v$ . All children of the same node are said to be *siblings* of each other. The *tree children* of a node  $u$  are its children that are tree nodes.

Let  $S$  be any finite set of *labels*. We say that the DAG  $N$  is *labeled in  $S$*  when its leaves are bijectively labeled by elements of  $S$ . Two DAGs  $N, N'$  labeled in  $S$  are *isomorphic*, in symbols  $N \cong N'$ , when they are isomorphic as directed graphs and the isomorphism preserves the leaves' labels.

We shall always assume, usually without any further notice, that the DAGs appearing in this paper are labeled in some set  $S$ , and we shall always identify, usually without any further notice either, each leaf of a DAG with its label in  $S$ .

A *path* on a DAG  $N = (V, E)$  is a sequence of nodes  $(v_0, v_1, \dots, v_k)$  such that  $(v_{i-1}, v_i) \in E$  for all  $i = 1, \dots, k$ ; such a path is a *cycle* if  $v_k = v_0$ . We call  $v_0$  the *origin* of the path,  $v_1, \dots, v_{k-1}$  its *intermediate nodes*, and  $v_k$  its *end*. The *length* of the path  $(v_0, v_1, \dots, v_k)$  is  $k$ , and it is *non-trivial* if  $k \geq 1$ . We denote by  $u \rightsquigarrow v$  any path with origin  $u$  and end  $v$ .

The relation  $\geq$  on  $V$  defined by

$$u \geq v \iff \text{there exists a path } u \rightsquigarrow v$$

is a partial order, called the *path ordering* on  $N$ . Whenever  $u \geq v$ , we shall say that  $v$  is a *descendant* of  $u$  and also that  $u$  is an *ancestor* of  $v$ .

### 3 Tripartitions

Let  $N = (V, E)$  be any DAG labeled in  $S$ . For every node  $u \in V$ :

- Let  $C(u) \subseteq S$  be the set of leaves that are descendants of  $u$ . We call  $C(u)$  the *cluster* of  $u$ .
- Let  $A(u) \subseteq C(u)$  be the set of leaves that are *strict descendants* of  $u$ : that is, those leaves  $s$  such that every path from a root of  $N$  to  $s$  contains the node  $u$ . We call  $A(u)$  the *strict cluster* of  $u$ .
- Let  $B(u) \subseteq C(u)$  be the set  $C(u) \setminus A(u)$  of leaves that are *non-strict descendants* of  $u$ : those leaves  $s$  that are descendants of  $u$ , but for which there exists some path from a root to  $s$  not containing the node  $u$ .
- Let  $C^c(u) \subseteq S$  be the set  $V_L \setminus C(u)$  of leaves that are not descendants of  $u$ .

A *phylogenetic tree* on a set  $S$  of taxa is a rooted tree with its leaves labeled bijectively in  $S$ , i.e., a rooted DAG labeled in  $S$  without hybrid nodes. Notice that, in a phylogenetic tree,  $C(u) = A(u)$  and  $B(u) = \emptyset$  for every node  $u$ . This property actually characterizes phylogenetic trees among all rooted DAGs.

Every arc  $e = (u, v)$  of a phylogenetic tree  $T = (V, E)$  on  $S$  defines a *bipartition* of  $S$

$$\pi^{(T)}(e) = (C(v), C^c(v)).$$

Let  $\pi(T)$  denote the set of all these bipartitions:

$$\pi(T) = \{\pi^{(T)}(e) \mid e \in E\}.$$

The *bipartition distance* [3, 18] between two phylogenetic trees  $T$  and  $T'$  on the same set  $S$  of taxa is defined as

$$d_\pi(T, T') = \frac{1}{2}(|\pi(T) \setminus \pi(T')| + |\pi(T') \setminus \pi(T)|).$$

The bipartition distance is a true distance for phylogenetic trees, in the sense that it satisfies the axioms of distances up to isomorphisms: for every phylogenetic trees  $T, T', T''$  on the same set  $S$  of taxa,

- (a) *Non-negativity*:  $d_\pi(T, T') \geq 0$
- (b) *Separation*:  $d_\pi(T, T') = 0$  if and only if  $T \cong T'$
- (c) *Symmetry*:  $d_\pi(T, T') = d_\pi(T', T)$
- (d) *Triangle inequality*:  $d_\pi(T, T') \leq d_\pi(T, T'') + d_\pi(T'', T')$

Phylogenetic networks [1] are usually defined as a subclass of DAGs that extend phylogenetic trees by allowing the existence of hybrid nodes representing recombination or lateral gene transfer events. In a non-tree DAG  $N$  it still makes sense to consider bipartitions  $\pi$  associated to arcs  $e = (u, v)$ ,

$$\pi^{(N)}(e) = (C(v), C^c(v)),$$

and then to define

$$\pi(N) = \{\pi^{(N)}(e) \mid e \in E\}.$$

and

$$d_\pi(N, N') = \frac{1}{2}(|\pi(N) \setminus \pi(N')| + |\pi(N') \setminus \pi(N)|).$$

As we shall see in Section 9, there even exist subclasses of non-tree DAGs where  $d_\pi$  is a distance.

But, to distinguish between strict and non-strict descendants gives more information about the topological relations between the hybrid nodes. This is the reason why, in the series of papers [11, 12, 13, 14, 15, 16, 17], Moret, Nakhleh, Warnow, and collaborators associate to each arc  $e = (u, v)$  of a DAG  $N$  labeled in  $S$ , the *tripartition* of  $S$

$$\theta^{(N)}(e) = (A(v), B(v), C^c(v)).$$

As an extra piece of information about the topology of the DAG, the leaves  $s$  in  $A(v)$  and  $B(v)$  can be weighted with the maximum number of hybrid nodes contained in a path from  $v$  to  $s$  (including  $v$  and  $s$  themselves). Therefore, we can distinguish between:

- The (*unweighted*) *tripartition*  $\theta^{(N)}(e)$  [16, 15, 14].
- The *B-weighted tripartition*  $\theta_B^{(N)}(e)$ , where the elements of  $B(v)$  are weighted as indicated [13].
- The *AB-weighted tripartition*  $\theta_{AB}^{(N)}(e)$ , where the elements of both  $A(v)$  and  $B(v)$  are weighted as indicated [11, 12].

For every type of tripartition  $\Upsilon = \theta, \theta_B, \theta_{AB}$ , let us denote by  $\Upsilon(N)$  the set of all these tripartitions of arcs of  $N$ :

$$\Upsilon(N) = \{\Upsilon^{(N)}(e) \mid e \in E\}.$$

The *tripartition distance relative to  $\Upsilon$*  between two DAGs  $N_1 = (V_1, E_1)$  and  $N_2 = (V_2, E_2)$  labeled in the same set  $S$  can be defined then, by analogy with the bipartition distance, by

$$d_{\Upsilon}(N_1, N_2) = \frac{1}{2}(|\Upsilon(N_1) \setminus \Upsilon(N_2)| + |\Upsilon(N_2) \setminus \Upsilon(N_1)|),$$

It is obvious that  $d_{\Upsilon}$  always satisfies the non-negativity, symmetry and triangle inequality axioms of distances on the class of all DAGs labeled in a fixed set  $S$ . As far as the separation axiom goes, notice that  $d_{\Upsilon}(N_1, N_2) = 0$  if and only if  $\Upsilon(N_1) = \Upsilon(N_2)$ . Therefore,  $d_{\Upsilon}$  satisfies the separation axiom on a certain subclass of DAGs if and only if  $\Upsilon(N)$  characterizes  $N$  up to isomorphism among all DAGs in this subclass. When these equivalent conditions hold, we shall say that the tripartition  $\Upsilon$  *satisfies the separation property* on the subclass of DAGs under consideration.

Notice that, for every DAGs  $N$  and  $N'$ ,

$$\theta_{AB}(N) = \theta_{AB}(N') \implies \theta_B(N) = \theta_B(N') \implies \theta(N) = \theta(N').$$

Then, the separation property for  $\theta$  implies the separation property for  $\theta_B$ , and the latter implies the separation property for  $\theta_{AB}$ .

In the papers by Moret, Nakhleh, Warnow, et al mentioned above, it is claimed that some of these tripartitions  $\Upsilon$  (and even further refinements of them) satisfy the separation property on some specific subclasses of DAGs. In the following sections we show that all these claims are incorrect, and then in Section 9 we show a subclass of phylogenetic networks where  $\theta$ , an even the bipartition  $\pi$ , satisfy the separation property.

To end this section, we want to mention that in the aforementioned papers, the  $\Upsilon$ -tripartition distance is actually not defined as above, but in a normalized version, which the authors call *error metric*:

$$m_{\Upsilon}(N_1, N_2) = \frac{1}{2} \left( \frac{|\Upsilon(N_1) \setminus \Upsilon(N_2)|}{|E_1|} + \frac{|\Upsilon(N_2) \setminus \Upsilon(N_1)|}{|E_2|} \right).$$

This is the function claimed to be a distance on some subclasses of DAGs in those papers. It is straightforward to notice that  $m_{\Upsilon}$  satisfies the separation axiom of distances on a subclass of DAGs if and only if  $d_{\Upsilon}$  does so, and therefore the counterexamples in the next sections also entail that  $m_{\Upsilon}$  neither satisfies the separation axiom when claimed.

But, contrary to what happens with  $d_{\Upsilon}$  and against what is claimed in several of the papers under review (see, for instance, the proof of [14, Thm. 3]), this normalized version  $m_{\Upsilon}$  need not satisfy the triangle inequality either, even on the subclasses of DAGs considered in those papers. For one reason, the failure of the separation property allows the existence of DAGs  $N = (V, E)$  and  $N' = (V', E')$  such that  $\Upsilon(N) = \Upsilon(N')$  but, say,  $|E| < |E'|$ . Then, given any other DAG  $N_0 = (V_0, E_0)$ ,

$$\begin{aligned} m_{\Upsilon}(N_0, N) &= \frac{1}{2} \left( \frac{|\Upsilon(N_0) \setminus \Upsilon(N)|}{|E_0|} + \frac{|\Upsilon(N) \setminus \Upsilon(N_0)|}{|E|} \right) \\ m_{\Upsilon}(N_0, N') &= \frac{1}{2} \left( \frac{|\Upsilon(N_0) \setminus \Upsilon(N')|}{|E_0|} + \frac{|\Upsilon(N') \setminus \Upsilon(N_0)|}{|E'|} \right) \\ &= \frac{1}{2} \left( \frac{|\Upsilon(N_0) \setminus \Upsilon(N)|}{|E_0|} + \frac{|\Upsilon(N) \setminus \Upsilon(N_0)|}{|E'|} \right) \\ m_{\Upsilon}(N', N) &= \frac{1}{2} \left( \frac{|\Upsilon(N') \setminus \Upsilon(N)|}{|E'|} + \frac{|\Upsilon(N) \setminus \Upsilon(N')|}{|E|} \right) = 0 \end{aligned}$$

and then  $|E| < |E'|$  implies that

$$\begin{aligned}
m_{\Upsilon}(N_0, N) &= \frac{1}{2} \left( \frac{|\Upsilon(N_0) \setminus \Upsilon(N)|}{|E_0|} + \frac{|\Upsilon(N) \setminus \Upsilon(N_0)|}{|E|} \right) \\
&> \frac{1}{2} \left( \frac{|\Upsilon(N_0) \setminus \Upsilon(N)|}{|E_0|} + \frac{|\Upsilon(N) \setminus \Upsilon(N_0)|}{|E'|} \right) + 0 \\
&= m_{\Upsilon}(N_0, N') + m_{\Upsilon}(N', N).
\end{aligned}$$

For two specific  $N = (V, E)$  and  $N' = (V', E')$  such that  $\Upsilon(N) = \Upsilon(N')$  but  $|E| \neq |E'|$ , precisely in the context of Thm. 3 in [14], see the DAGs  $N_9$  and  $N_{10}$  depicted in Fig. 8 in the Appendix<sup>1</sup>. They are reduced reconstructible phylogenetic networks (see Section 8 for the precise meaning of these words), the subclass where that theorem claims that  $m_{\theta}$  satisfies the triangle inequality, and they have the same  $AB$ -weighted tripartitions (see Table 7 also in the Appendix) but different numbers of arcs. This shows that  $m_{\theta_{AB}}$  does not satisfy the triangle inequality on the subclass of reduced reconstructible phylogenetic networks.

Anyway, the failure of the triangle inequality is easily solved for instance using  $d_{\Upsilon}$  instead of  $m_{\Upsilon}$ . The failure of the separation axiom is deeper, as it reflects the impossibility of discriminating the phylogenetic networks under consideration using only information on their tripartitions.

## 4 The first error metric

The error metric for phylogenetic networks is introduced by Moret, Nakhleh, and Warnow in the Technical Report [13]. In it, a *phylogenetic network* on a set  $S$  of taxa is defined as a rooted DAG  $N$  labeled in  $S$  satisfying the following conditions:

- (4.1) The in-degree and out-degree of each node is 0, 1, or 2, and no node has its in-degree equal to its out-degree.
- (4.2) If a node has two children, at least one of them is a tree node.
- (4.3) *Weak time consistency*:
  - (4.3.a) If  $u_1$  and  $u_2$  are the parents of a hybrid node  $u$ , then there do not exist paths  $u_1 \rightsquigarrow u_2$  or  $u_2 \rightsquigarrow u_1$ .
  - (4.3.b) If  $u_1$  and  $u_2$  are the parents of a hybrid node  $u$ , and  $v_1$  and  $v_2$  are the parents of a hybrid node  $v$ , and there exists a path  $u_1 \rightsquigarrow v_1$ , then there do not exist paths  $v_2 \rightsquigarrow u_1$  or  $v_2 \rightsquigarrow u_2$ .

Notice that in these phylogenetic networks, a hybrid node can have a hybrid child. This would correspond to a hybrid node that hybridizes before undergoing a speciation event, a scenario that, the authors say, “almost never arises in reality.”

Condition (4.3) is a first, weak version of a constant property of Nakhleh-Warnow-et al’s phylogenetic networks: time consistency. Roughly speaking, this property aims at assigning times to the nodes of the network in a way that strictly increases on tree arcs and so that the parents of a hybrid node coexist in time; we shall discuss this topic further in the next section. In the weak version recalled in this section, and which does not entail in general such a timing (cf. Prop. 1 in the next section), this restriction is simply imposed by not

---

<sup>1</sup>To make easier reading this paper, we gather all large figures depicting phylogenetic networks, as well as the tables giving their tripartitions, in an Appendix at the end of the paper.

allowing a node to hybridize with its descendants, and by forbidding the ancestors of a parent of a hybrid node to hybridize with the descendants of the other parent.

A phylogenetic network is said to be of *class I* when each hybrid node has at least one parent that is a tree node.

Then, it is claimed in [13, Thm. 4] that  $\theta_B$  satisfies the separation property on the subclass of all class I phylogenetic networks. This claim is untrue, because there exist pairs of non-isomorphic class I phylogenetic networks with the same sets of  $B$ -weighted (even  $AB$ -weighted) tripartitions.

Consider for instance the phylogenetic networks  $N_1$  and  $N_2$  labeled in  $\{1, \dots, 5\}$  depicted in Fig. 3 in the Appendix. It is easy to check that these two DAGs satisfy conditions (4.1) to (4.3) above and that they are of class I. Now, Table 1 displays the  $AB$ -weighted tripartitions of these networks induced by their arcs. A simple inspection of this table shows that  $\theta_{AB}(N_1) = \theta_{AB}(N_2)$ . But it is clear that  $N_1 \not\cong N_2$ .

It is interesting to point out that  $\theta_B$ , and even  $\theta$ , satisfy the separation property if we moreover forbid the “improbable” event of two consecutive hybridizations: i.e., if we impose not only that some child of a tree node is a tree node (condition (4.2)), but also that the only child of an internal hybrid node is a tree node, a condition that would be imposed in later versions (see conditions (5.2), (6.2), and (8.2)). We devote Section 9 to prove this fact.

## 5 The second version: introducing strong time consistency

A new version of the tripartition metric is presented in the Technical Report [11]. This is the metric used, for instance, in the paper [17]. The main difference between the proposal in this new Technical Report and the previous one [13] is the refinement of the notion of phylogenetic network, by distinguishing between *model* and *reconstructible* phylogenetic networks and strengthening the time compatibility and class I conditions.

In [11], a *model phylogenetic network* on a set  $S$  of taxa is defined as a rooted DAG  $N$  labeled in  $S$  satisfying the following conditions:

- (5.1) The root and all internal tree nodes have out-degree 2. All hybrid nodes have out-degree 1, and they can only have in-degree 2 (allo-polyploid hybrid nodes) or 1 (auto-polyploid hybrid nodes).<sup>2</sup>
- (5.2) The child of a hybrid node is always a tree node.
- (5.3) *Strong time consistency*: Let  $x, y$  be any two nodes for which there exists a sequence of nodes  $(v_0, v_1, \dots, v_k)$  with  $v_0 = x$  and  $v_k = y$  such that:

- for every  $i = 0, \dots, k-1$ , either  $(v_i, v_{i+1})$  is an arc of  $N$ , or  $(v_{i+1}, v_i)$  is a network arc of  $N$ ;
- at least one pair  $(v_i, v_{i+1})$  is a tree arc of  $N$ ;

(that is,  $(v_0, \dots, v_k)$  is a path from  $x$  to  $y$  containing some tree arc of  $N$ , in the graph  $N^*$  obtained from  $N$  by adding the inverses of all network arcs). Then,  $x$  and  $y$  cannot have a hybrid child in common.

---

<sup>2</sup>In our classification of nodes in DAGs in Section 2, these auto-polyploid hybrid nodes, with in-degree and out-degree equal to 1, would actually be considered tree nodes.

This new notion of time consistency generalizes the one given in the previous version, and, as we shall see in a minute, it captures the notion of timing mentioned therein. This timing is given by a *temporal representation* in the sense of Baroni, Sempel and Steel [2]: a mapping  $\tau : V \rightarrow \mathbb{N}$  such that:

- (a) if  $r$  is the root of  $N$ , then  $\tau(r) = 0$ ;
- (b) if  $(u, v) \in E_T$ , then  $\tau(u) < \tau(v)$ ;
- (c) if  $(u, v) \in E_N$ , then  $\tau(u) = \tau(v)$ .

Baroni, Sempel and Steel prove in *loc. cit.* the equivalence between the existence of such a temporal representation and the fact that a certain quotient graph (essentially obtained by identifying hybrid nodes with their parents) of the network is acyclic. Since none of the papers on tripartitions we are discussing provides a formal proof of the fact that condition (5.3) above is equivalent to the existence of a temporal representation, and for the sake of completeness, we provide such a proof here.

**Proposition 1.** *Let  $N = (V, E)$  be a rooted DAG, let  $E_T$  and  $E_N$  be its sets of tree and network arcs, respectively, and let  $N^* = (V, E^*)$  be the directed graph with the same set  $V$  of nodes as  $N$  and set of arcs  $E^* = E \cup E_N^{-1}$ . The following conditions are equivalent:*

- (i)  $N$  is strongly time consistent.
- (ii)  $N^*$  does not have any cycle containing some tree arc of  $N$ .
- (iii)  $N$  admits a temporal representation.

*Proof.* (i) $\implies$ (ii) To begin with, notice that if  $N^*$  has cycles containing tree arcs of  $N$ , then it has a minimal<sup>3</sup> such cycle. Indeed, if  $N^*$  has cycles containing tree arcs, then it will contain one of shortest length, say

$$(v_0, v_1, \dots, v_k, v_0).$$

If it is not minimal, then  $v_i = v_j$  for some  $0 \leq i < j \leq k$  (actually  $j - i \geq 2$ , because  $N$  does not contain loops), and hence we have two strictly shorter cycles in  $N^*$ ,

$$(v_i, v_{i+1}, \dots, v_{j-1}, v_j = v_i) \text{ and } (v_0, v_1, \dots, v_{i-1}, v_i = v_j, v_{j+1}, \dots, v_k, v_0),$$

and at least one of them contains the tree arc that belonged to the original cycle, which leads to a contradiction.

Assume now that  $N$  satisfies the strong time consistency condition and that  $N^*$  has a minimal cycle

$$(v_0, v_1, \dots, v_k, v_0)$$

containing some tree arc of  $N$ : without any loss of generality, we shall assume that  $(v_k, v_0)$  is a tree arc of  $N$ , and in particular that  $v_0$  is a tree node.

Since  $N$  is acyclic, this cycle must contain some arc in  $E_N^{-1}$ . Let  $i \in \{1, \dots, k-1\}$  be the first index such that  $(v_i, v_{i+1}) \in E_N^{-1}$ : in particular,  $v_{i+1}$  is one of the parents in  $N$  of the hybrid node  $v_i$  (notice that  $i \neq 0$ , because  $v_0$  is a tree node of  $N$ , and that  $i \neq k$  because  $(v_k, v_0)$  is a tree arc of  $N$ ). Then the arc  $(v_{i-1}, v_i)$  is a network arc of  $N$ , and since the considered cycle is minimal,  $v_{i-1}$  must be different from  $v_{i+1}$ .

---

<sup>3</sup>By a *minimal cycle*  $(v_0, v_1, \dots, v_k, v_0)$  we mean a cycle such that the nodes  $v_0, v_1, \dots, v_k$  are pairwise different.



But in this case the sequence of nodes

$$(v_{i+1}, \dots, v_k, v_0, v_1, \dots, v_{i-1})$$

is a path in  $N^*$  containing at least one tree arc and connecting two parents of a hybrid node of  $N$ , which contradicts the strong time consistency condition. This shows that  $N^*$  cannot contain any minimal cycle containing some tree arc of  $N$ .

(ii) $\implies$ (iii) If  $N^*$  does not have any cycle containing some tree arc of  $N$ , then the mapping

$$\tau : V \rightarrow \mathbb{N}$$

that sends each  $v \in V$  to the maximum number of tree arcs in a path from the root  $r$  to  $v$  in  $N^*$  is well defined, and it clearly satisfies conditions (a) to (c) in the statement.

(iii) $\implies$ (i) Assume that a mapping  $\tau$  as described in (iii) exists. Then, if  $(v_0, \dots, v_k)$  is a path in  $N^*$  containing some tree arc of  $N$ , we have that  $\tau(v_0) < \tau(v_k)$ , and then  $v_0$  and  $v_k$  cannot be the parents of a hybrid node  $u$ , because the latter would imply that  $\tau(v_0) = \tau(u) = \tau(v_k)$ .  $\square$   $\square$

In *reconstructible phylogenetic networks* the previous conditions are relaxed: tree nodes can have any out-degree greater than 1; hybrid nodes can have any in-degree greater than 1 and any out-degree greater than 0 (in particular, auto-polyploid hybrid nodes are forbidden); a hybrid node can have hybrid children; and the strong time consistency need not hold any longer.

Now, two nodes  $u, v$  of a (model or reconstructible) phylogenetic network are said to be *convergent* when they satisfy the following condition:

For every leaf  $s$  and for every  $k \geq 0$ , there exists a path  $u \rightsquigarrow s$  containing  $k$  hybrid nodes if and only if there exists a path  $v \rightsquigarrow s$  containing  $k$  hybrid nodes.

A phylogenetic network is said to be *of class I* if it does not contain any pair of convergent nodes.

Then, it is claimed in [11, Thm. 4] that  $\theta_{AB}$  satisfies the separation property on this new class I of phylogenetic networks. It is again incorrect, because there still exist pairs of non-isomorphic class I model phylogenetic networks with  $AB$ -weighted error rate 0. Consider for instance the model phylogenetic networks  $N_3$  and  $N_4$  labeled in  $\{1, \dots, 13\}$  depicted in Fig. 4 in the Appendix, which are suitable modifications of those given in Fig. 3 to cope with the new restrictions on model phylogenetic networks (much simpler examples exist involving reconstructible networks containing, for instance, out-degree 3 tree nodes: consider the networks  $N_9$  and  $N_{10}$  shown in Fig. 8).

These networks have no pair of convergent nodes (as it can be checked in Table 2) and they are clearly non-isomorphic. Now, Table 2 displays the tripartitions of these networks induced by their arcs and, again, a simple inspection of this table shows that  $\theta_{AB}(N_3) = \theta_{AB}(N_4)$ .

## 6 The third version: introducing the reticulation scenarios

In the Technical Reports [12, 16] the authors present a third version of the tripartition metric, now including a substantial change in the definition of the metric itself.

The definition of phylogenetic network in these papers is that of model phylogenetic network in the previous section with a small modification in the time compatibility condition. More specifically, a *phylogenetic network* on a set  $S$  of labels is a rooted DAG  $N$  labeled in  $S$  satisfying the following conditions:

- (6.1) The root and all internal tree nodes have out-degree 2. All hybrid nodes have out-degree 1, and they can only have in-degree 2 or (in [12]) 1.
- (6.2) The child of a hybrid node is always a tree node.
- (6.3) *Time consistency*: Let  $x, y$  be any two nodes for which there exists a sequence of paths  $(P_0, P_1, \dots, P_k)$  in  $N$  such that
- $x$  is the origin of  $P_0$  and  $y$  is the end of  $P_k$ ;
  - each path contains some tree arc;
  - for every  $i = 0, \dots, k-1$ , the end of  $P_i$  and the origin of  $P_{i+1}$  are the parents of a hybrid node.

Then,  $x$  and  $y$  cannot have a hybrid child in common.

This time consistency condition (6.3) is actually weaker than the strong time consistency (5.3). For instance, Fig. 1 shows two situations where condition (6.3) allows the nodes  $u$  and  $v$  to hybridize, while condition (5.3) forbids it. Notice nevertheless that, under condition (6.2), time consistency (6.3) becomes equivalent to strong time consistency (5.3) if we simply ask *at least one of the paths*  $P_i$  (instead of all of them) to contain some tree arc.

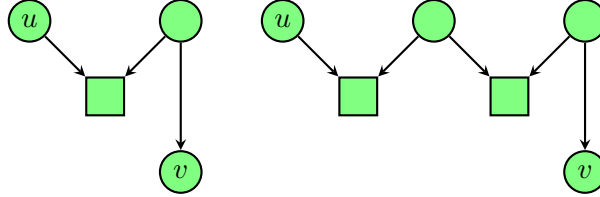


Figure 1: Time compatibility allows the nodes  $u$  and  $v$  in these graphs to hybridize, while strong time compatibility does not.

In these papers the authors do not consider any class I of phylogenetic networks, but instead they refine the error rate by adding some extra information to tripartitions induced by network arcs. Namely, they define the *reticulation scenario*  $RS(v)$  of a hybrid node  $v$  with parents  $u_1, u_2$  as the set of clusters of its parents:

$$RS(v) = \{C(u_1), C(u_2)\}.$$

Then, the data the authors<sup>4</sup> consider on arcs are:

- If  $e$  is a tree arc, then  $\Psi^{(N)}(e) = \theta_{AB}^{(N)}(e)$ ;
- if  $e$  is a network arc with head  $v$ , then  $\Psi^{(N)}(e) = (\theta^{(N)}(e), RS(v))$ .

We shall call  $\Psi^{(N)}(e)$  the *enriched (AB-weighted) tripartition* of  $N$  associated to  $e$ . Notice that  $\Psi^{(N)}(e)$  still depends always only on  $e$ 's head.

Let us denote by  $\Psi(N)$  the set of all these enriched tripartitions:

$$\Psi(N) = \{\Psi^{(N)}(e) \mid e \in E\}.$$

---

<sup>4</sup>Actually, the tripartitions they use are *AB-weighted* in [12] and *unweighted* in [16]. For the sake of generality, in this section we shall use  $\theta_{AB}$ .

Then, in the papers quoted above, these sets  $\Psi(N)$  are used to define a metric  $m_\Psi$  by means of a formula similar to that of  $m_\Upsilon$  recalled in Section 3.

It is claimed in [12, §5.4] and [16, §5] that  $\Psi$  satisfies the separation property on the class of all phylogenetic networks, in the sense that if  $\Psi(N_1) = \Psi(N_2)$  then  $N_1 \cong N_2$ , for every phylogenetic networks  $N_1$  and  $N_2$  labeled in the same set  $S$ . And again, it is untrue, as there exist pairs of non-isomorphic phylogenetic networks with the same sets of enriched tripartitions. For instance, the phylogenetic networks  $N_3$  and  $N_4$  labeled in  $\{1, \dots, 13\}$ , used already in the previous section and depicted in Fig. 4. They have the same hybrid nodes, and we have already seen in Table 2 that they have the same sets of tripartitions induced by tree arcs as well as the same sets of tripartitions induced by network arcs. Table 3 shows that their hybrid nodes have the same reticulation scenarios. From these two tables, one can easily check that  $\Psi(N_3) = \Psi(N_4)$ .

## 7 Nakhleh's thesis: introducing the tree-sibling condition

In his PhD Thesis [15], Nakhleh uses the enriched tripartitions  $\Psi$  (actually, he uses un-weighted tripartitions, but for the sake of generality we shall still use them *AB-weighted*), but he restricts the subclass of phylogenetic networks where  $\Psi$  is stated to satisfy the separation property.

The definition of *model* phylogenetic network in this work is that of phylogenetic network given in the last section (conditions (6.1), (6.2) and (6.3)), while in *reconstructible* networks these conditions are relaxed as in Section 5.

Then he defines a phylogenetic network to be of *class I* when every hybrid node has at least one sibling that is a tree node. We shall say henceforth that such a phylogenetic network satisfies the *tree-sibling condition*, or simply that it is *tree-sibling*, to distinguish these networks from previous class I networks defined through the absence of convergent pairs: notice that, for instance, phylogenetic networks  $N_3$  and  $N_4$  in Fig. 4 have no pair of convergent nodes, but they are not tree-sibling, while networks  $N_5$  and  $N_6$  in Fig. 5 in the Appendix are tree-sibling, but have pairs of convergent nodes (see Table 4). The phylogenetic networks used in [9, 10] (obtained by adding network arcs to a phylogenetic tree by repeating the following procedure: choose pairs of arcs  $(u_1, v_1)$  and  $(u_2, v_2)$  in the tree; split the first into  $(u_1, w_1)$  and  $(w_1, v_1)$ , with  $w_1$  a new (tree) node; split the second one into  $(u_2, w_2)$  and  $(w_2, v_2)$ , with  $w_2$  a new (hybrid) node; finally, add a new arc  $(w_1, w_2)$ ) are tree-sibling.

Nakhleh claims [15, Thm. 4 in Ch. 6] that  $\Psi$  satisfies the separation property on the subclass of all tree-sibling phylogenetic networks. It is false, as there exist pairs of non-isomorphic tree-sibling model phylogenetic networks with the same sets of enriched tripartitions. Consider for instance the phylogenetic networks  $N_5$  and  $N_6$  labeled in  $\{1, \dots, 10\}$  depicted in Fig. 5 in the Appendix. They are model phylogenetic networks (they even satisfy the strong time consistency condition (5.3) instead of the time consistency condition (6.3)), they satisfy the tree-sibling condition, and they are clearly non-isomorphic.

Table 4 displays the tripartitions of these networks induced by their arcs, and Table 5 gives the reticulation scenarios of their hybrid nodes (which are the same in both networks). From these two tables, one can easily check that  $\Psi(N_5) = \Psi(N_6)$ .

## 8 Tripartitions do not distinguish distinguishable networks

In the final paper [14] of this series, Moret, Nakhleh, Warnow et al assert that their tripartitions can be used to distinguish networks up to a certain notion of reduction that we recall below, from where they deduce that  $\theta$  satisfies the separation property on a very restricted subclass of phylogenetic networks.

The notion of *model phylogenetic network* in this paper is exactly that of the last two sections. As far as reconstructible phylogenetic networks goes, they do not relax them as much as in previous papers. More specifically, a *reconstructible phylogenetic network* on a set  $S$  of labels is defined as a rooted DAG labeled in  $S$  satisfying the following conditions:

- (8.1) The root and all internal tree nodes can have any out-degree greater than 1. All hybrid nodes have out-degree 1, but they can have any in-degree greater than 1.
- (8.2) The child of a hybrid node is always a tree node.
- (8.3) Time consistency property (6.3).

A subset  $U$  of internal nodes of  $V$  is said to be *convergent* when it has at least two elements, and all nodes in it have exactly the same cluster (contrary to previous versions, no condition on the number of hybrid nodes in the paths to descendant leaves is required). The removal of convergent sets is the basis of the following *reduction procedure*:

- (0) Replace every clade by a new “symbolic leaf” labeled with the names of all leaves in it.
- (1) For every maximal convergent set  $U$ , remove all nodes in paths from nodes in  $U$  to (symbolic) leaves, including the node in  $U$  but keeping the leaf. For every node  $x$  that is the tail of an arc whose head  $v$  has been removed, and for every leaf  $s$  in the cluster of  $v$ , add a new arc  $(x, s)$ .  
(The resulting network contains no convergent set of nodes, because this step does not change the clusters of the surviving nodes.)
- (2) Append to every symbolic leaf representing a clade the corresponding clade, with an arc from the symbolic leaf to the root of the clade.
- (3) Replace every path of length greater than 1 with all its intermediate nodes of in- and out-degree equal to 1 by a single arc from its origin to its end.  
(In particular, if a symbolic leaf turns out to have only one parent, then it is removed and the root of the corresponding clade is appended to its first ancestor with out-degree different from 1.)

The output of this procedure applied to a reconstructible phylogenetic network  $N$  is a DAG  $R(N)$  labeled in  $S$ . The network  $R(N)$  is called the *reduced version* of  $N$ . Two networks  $N_1$  and  $N_2$  are said to be *indistinguishable* when they have isomorphic reduced versions, that is, when  $R(N_1) \cong R(N_2)$ .

Since every hybrid node in  $N$  and its only child form a convergent set, they are removed in step (1) together with all their descendants until the clades’ symbolic leaves. On the other hand, in  $R(N)$  the symbolic leaves may have more than one parent, and then they are the only possible hybrid nodes in  $R(N)$ . So, in particular, no hybrid node in  $R(N)$  is a descendant of another hybrid node. Moreover, since all convergent sets and all nodes with in- and out-degree 1 in  $N$  are removed, the only possible convergent sets in  $R(N)$  consist

of a hybrid node and its only child (that is, a symbolic leaf with more than one parent and the root of the corresponding clade).

We want to point out that the reduced version of a reconstructible phylogenetic network need not be a reconstructible phylogenetic network. Consider for instance the simple network  $N$  in Fig. 2 below. The nodes  $a, b, r$  form a convergent set, and therefore they are removed in the reduction process. Then, the reduced version of  $N$  is a non connected DAG consisting of four arcs with heads the leaves of  $N$ . As another example, consider the reconstructible phylogenetic network  $N_8$  in Fig. 6 in the Appendix: its reduced version, which is shown in Fig. 7, does not satisfy the time consistency property. This shows that reduced versions need not be time consistent.

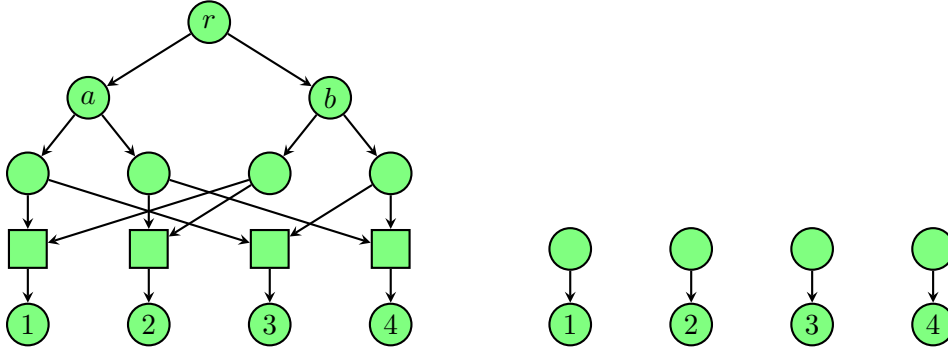


Figure 2: A model phylogenetic network  $N$  (left) and its reduced version (right)

It is claimed in [14, Lem. 1] that if  $\theta(N_1) = \theta(N_2)$ , then  $N_1$  and  $N_2$  are indistinguishable (the converse does not hold in general, because the reduction process may remove parts with different topologies that yield differences in the sets of tripartitions). This is not true. Consider for instance the model phylogenetic networks  $N_7$  and  $N_8$  depicted in Fig. 6 in the Appendix. Table 6 displays the tripartitions of these networks induced by their arcs, showing that the sets of tripartitions are the same (we give actually the  $AB$ -weighted tripartitions, just to show that their claim is still false if we replace  $\theta$  by  $\theta_{AB}$ ). Fig. 7 shows the reduced versions of these networks, and they are clearly non isomorphic.

The authors also claim [14, Thm. 3] that  $\theta$  satisfies the separation property on the subclass of reduced reconstructible phylogenetic networks (that is, of reconstructible phylogenetic networks that remain untouched under the reduction procedure). This is also wrong. Consider for instance the networks  $N_9$  and  $N_{10}$  depicted in Fig. 8 in the Appendix. They are reconstructible phylogenetic networks, and they are reduced because the only convergent sets they contain consist of a hybrid node and a leaf that is its only child, and therefore the application of the reduction procedure leaves them untouched. Table 7 shows that these two networks have the same sets of  $AB$ -weighted tripartitions, but they are clearly non-isomorphic.

## 9 Tree-child phylogenetic networks

We have seen in previous sections that neither the lack of convergent pairs of nodes, nor the tree-sibling condition or even the property of being reduced, in all cases combined with the strong time consistency condition, guarantee the separation property for  $\theta_{AB}$ . In this section we introduce a stronger condition that, by itself, does not guarantee this separation

property either, but that combined with condition (4.3.a) makes  $d_\theta$  satisfy the separation property. Even more, it makes *bipartitions*  $\pi$  satisfy the separation property.

We shall say that a DAG satisfies the *tree-child condition*, or simply that it is *tree-child*, when every node other than a leaf has at least one tree child. A *tree-child phylogenetic network* is a rooted tree-child DAG with no tree node of out-degree 1 and all hybrid nodes of out-degree exactly 1 (and any in-degree greater than 1). So, tree-child phylogenetic networks can be understood as models of reticulated evolution where:

- The tree nodes represent species.
- The hybrid nodes represent recombination or lateral gene transfer events that yield the species corresponding to their single tree child.
- Every species other than the extant ones, represented by the leaves, have some descendant through mutation.

The (even enriched) tripartitions introduced so far do not satisfy the separation property on the subclass of *all* tree-child phylogenetic networks. Consider for instance the networks  $N_{11}$  and  $N_{12}$  depicted in Fig. 9 in the Appendix. Table 8 provides the  $AB$ -weighted tripartitions and reticulation scenarios of their arcs, showing that these networks cannot be distinguished using this information.

But if we add to the tree-child condition the weakest form of time consistency, namely condition (4.3.a) (two parents of a hybrid node cannot be connected by a path), then  $\theta$  satisfies the separation property on the resulting subclass of phylogenetic networks. Actually, it turns out that the bare sets of clusters of nodes (without distinguishing between strict and non-strict descendants, and without taking into account the numbers of hybrid nodes in paths to leaves) are enough to characterize these phylogenetic networks up to isomorphism (Thm. 8), which entails that  $\pi$  satisfies the separation property, just as in phylogenetic trees.

To prove these facts, we need to establish some preliminary definitions and results.

We shall denote henceforth by  $C^{(N)}(v)$  the cluster of a node  $v$  in a DAG  $N$  labeled in  $S$ , to emphasize the network. For every DAG  $N = (V, E)$ , let  $C(N)$  denote the set of clusters of its nodes:

$$C(N) = \{C^{(N)}(v) \mid v \in V\}.$$

A *tree path* in a DAG  $N$  is a path consisting entirely of tree arcs. A node  $v$  is a *tree descendant* of a node  $u$  when there exists a tree path from  $u$  to  $v$ .

**Lemma 2.** *Every node  $v$  in a tree-child phylogenetic network has some tree descendant leaf.*

*Proof.* If  $v$  is not already a leaf, we can construct a tree path starting in  $v$  by successively taking tree children, and this path will end in a leaf that will be a tree descendant of  $v$ . □ □

**Lemma 3.** *Let  $u \rightsquigarrow v$  be a tree path in a DAG  $N$ . Then, for every other path  $w \rightsquigarrow v$  ending in  $v$ , either it contains  $u \rightsquigarrow v$  or  $u \rightsquigarrow v$  contains  $w \rightsquigarrow v$ .*

*Proof.* Let  $(u, v_1, \dots, v_{k-1}, v)$  be the tree path  $u \rightsquigarrow v$  in the statement: to simplify the notations, we call  $v_0 = u$  and  $v_k = v$ . Let now  $w \rightsquigarrow v$  be any other path ending in  $v$  and let  $v_j$  be the first node in the path  $u \rightsquigarrow v$  such that  $(v_j, \dots, v_k)$  is contained in  $w \rightsquigarrow v$ . If  $j \neq 0$ , then  $v_j$  has only one parent, and therefore it must happen that either  $v_j = w$  or  $v_j$ 's parent

in the path  $w \rightsquigarrow v$  is also its parent in the path  $u \rightsquigarrow v$ , and in particular it also belongs to this path, which contradicts the assumption on  $v_j$ .

This proves that either  $v_j = w$ , in which case the path  $u \rightsquigarrow v$  contains the path  $w \rightsquigarrow v$ , or  $v_j = u$ , in which case the path  $w \rightsquigarrow v$  contains the path  $u \rightsquigarrow v$ .  $\square$   $\square$

**Corollary 4.** *If  $v$  is a tree descendant of  $u$  in a DAG  $N$ , then  $v \in A(u)$  and the path  $u \rightsquigarrow v$  is unique.*

*Proof.* If  $v$  is a tree descendant of  $u$ , then there exists a tree path  $u \rightsquigarrow v$ . Then, by the previous lemma, every path from a root  $r$  to  $v$  must contain this path  $u \rightsquigarrow v$ , which shows that  $v$  is a strict descendant of  $u$ , and every other path from  $u$  to  $v$  must contain (and hence be equal to) this path  $u \rightsquigarrow v$ .  $\square$   $\square$

**Lemma 5.** *Let  $N = (V, E)$  be a tree-child phylogenetic network satisfying condition (4.3.a). For every nodes  $u, v \in V$ , the following conditions are equivalent:*

- (i)  $C^{(N)}(u) = C^{(N)}(v)$
- (ii)  $u = v$  or  $\{u, v\}$  are a hybrid node and its only child.

*Proof.* The implication (ii) $\implies$ (i) is obvious. As far as the implication (i) $\implies$ (ii) goes, assume that  $C^{(N)}(u) = C^{(N)}(v)$  and that  $u \neq v$ . If  $s$  is a tree descendant leaf of  $u$ , then  $s \in C^{(N)}(v)$ , and hence, by Lemma 3, either  $u$  belongs to the path  $v \rightsquigarrow s$  or  $v$  belongs to the path  $u \rightsquigarrow s$ . Therefore,  $u$  and  $v$  are connected by a path. To fix ideas, assume that there exists a path  $u \rightsquigarrow v$ . Then we must distinguish three cases:

- If  $u$  is a tree node and it has some tree child  $w$  outside the path  $u \rightsquigarrow v$ , then every tree descendant leaf  $s$  of  $w$  is a tree descendant leaf of  $u$  and hence, since  $C^{(N)}(u) = C^{(N)}(v)$ , a descendant of  $v$ . By the uniqueness of the path  $u \rightsquigarrow s$  (Corollary 4), the tree path  $u \rightsquigarrow s$  must be equal to the concatenation of the path  $u \rightsquigarrow v$  and the path  $v \rightsquigarrow s$ , but these paths are different because their first arcs are different. This yields a contradiction.
- If  $u$  is a tree node and all its children outside the path  $u \rightsquigarrow v$  are hybrid, take one such hybrid child  $w$ . Let  $s$  be any tree descendant leaf of  $w$ . Then  $s \in C^{(N)}(u) = C^{(N)}(v)$  and therefore there exists a path  $v \rightsquigarrow s$ . Now, by Corollary 4, the path  $w \rightsquigarrow s$  is unique, and therefore the path  $u \rightsquigarrow s$  is also unique. Indeed, given any path  $u \rightsquigarrow s$ , by Lemma 3 and since  $u$  cannot be a descendant of  $w$ , the path  $w \rightsquigarrow s$  must be contained in this path  $u \rightsquigarrow s$ , and since no other parent of  $w$  is a descendant of  $u$  by (4.3.a), the only possibility is that the first arc of this path is  $(u, w)$  and then the rest of the path is the tree path  $w \rightsquigarrow s$ .

This means that the path obtained by concatenating the paths  $u \rightsquigarrow v$  and  $v \rightsquigarrow s$  must be equal to the path  $u \rightsquigarrow s$  through  $w$ , but, again, these paths are different because their first arcs are different. This yields again a contradiction.

- If  $u$  is a hybrid node and  $u'$  is its unique tree child, then  $C^{(N)}(u') = C^{(N)}(u) = C^{(N)}(v)$  and  $u'$  must be the first node after  $u$  in the path  $u \rightsquigarrow v$ . This yields a path  $u' \rightsquigarrow v$  with  $C^{(N)}(u') = C^{(N)}(v)$  and  $u'$  a tree node. The last two points have shown that the assumption that  $u' \neq v$  leads to a contradiction, while  $u' = v$  is clearly possible.

In summary, the only situation that does not lead to a contradiction is when  $u$  is a hybrid node and  $v$  its only child.  $\square$   $\square$

**Lemma 6.** *Let  $N = (V, E)$  be a tree-child phylogenetic network satisfying condition (4.3.a). For every nodes  $u, v \in V$  such that  $C^{(N)}(u) \neq C^{(N)}(v)$ , the following two conditions are equivalent:*

(i) *There is a non-trivial path  $u \rightsquigarrow v$ .*

(ii)  $C^{(N)}(v) \subsetneq C^{(N)}(u)$ .

*Proof.* The implication (i) $\implies$ (ii) is straightforward: if  $v$  is a descendant of  $u$ , then  $C^{(N)}(v) \subseteq C^{(N)}(u)$ , and by assumption these clusters are different.

As far as the converse implication goes, assume that  $C^{(N)}(v) \subsetneq C^{(N)}(u)$ , and let  $s$  be a tree descendant leaf of  $v$ . Then  $s \in C^{(N)}(u)$  and Lemma 3 entails that  $u$  and  $v$  are connected by a path. Since clusters decrease with paths, this path must be  $u \rightsquigarrow v$ .  $\square \square$

Now, given a tree-child phylogenetic network  $N = (V, E)$  satisfying condition (4.3.a), its *contracted version* is the DAG  $N' = (V', E')$  obtained from  $N$  by contracting into one node each pair of nodes consisting of a hybrid node and its only child: more specifically, for every hybrid node  $u$ , if  $\bar{u}$  is its only child and  $\bar{u}_1, \dots, \bar{u}_k$  the children of  $\bar{u}$ , then we remove this node  $\bar{u}$  and the arcs incident to it, and we replace the latter by new arcs  $(u, \bar{u}_1), \dots, (u, \bar{u}_k)$ . In this way, we understand  $V'$  as a subset of  $V$ , consisting of all nodes of  $N$  except the children of hybrid nodes.

It is clear that  $C^{(N)}(v) = C^{(N')}(v)$  for every node  $v \in V'$ , and hence that  $C(N) = C(N')$ . Moreover,  $C^{(N')}(u) \neq C^{(N')}(v)$  if  $u \neq v$ , because each pair of nodes in  $N$  with the same cluster has been contracted to a single node in  $N'$ . In particular, the mapping

$$\begin{aligned} C^{(N')} : V' &\rightarrow C(N') \\ v &\mapsto C^{(N')}(v) \end{aligned}$$

is bijective.

On the other hand, for every  $u, v \in V'$ , there exists a path  $u \rightsquigarrow v$  in  $N$  if and only if there exists a path  $u \rightsquigarrow v$  in  $N'$ . Therefore, from Lemma 6 we deduce that there exists a path  $u \rightsquigarrow v$  in  $N'$  if and only if  $C^{(N')}(v) \subseteq C^{(N')}(u)$ : that is, the inclusion of clusters in  $N'$  captures exactly the path ordering in  $N'$ , which is the restriction to  $V'$  of the path ordering on  $N$ .

**Lemma 7.** *Let  $N$  be a tree-child phylogenetic network satisfying condition (4.3.a), and let  $N' = (V', E')$  be its contracted version. For every  $u \in V'$ , let*

$$M_u = \{w \in V' \mid C^{(N')}(w) \subsetneq C^{(N')}(u)\}.$$

*Then, the maximal elements of  $M_u$  with respect to the path ordering on  $N'$  are exactly the children of  $u$  in  $N'$ .*

*Proof.* If  $u$  is a leaf, then  $C^{(N')}(u) = \{u\}$  and  $M_u = \emptyset$ , and the thesis of the statement clearly holds. So, assume that  $u$  is not a leaf. Then, every descendant of  $u$  is in  $M_u$  and therefore  $M_u$  is non-empty.

Since  $M_u$  is finite, it has maximal elements. Let  $v$  be any such a maximal element. Since  $C^{(N')}(v) \subsetneq C^{(N')}(u)$ , there exists a non trivial path  $u \rightsquigarrow v$  in  $N'$ . If this path passes through some other node  $w$ , then  $C^{(N')}(v) \subsetneq C^{(N')}(w) \subsetneq C^{(N')}(u)$ , against the assumption that  $v$  is maximal in  $M_u$ . Therefore, the path  $u \rightsquigarrow v$  has length 1 and  $v$  is a child of  $u$ .

Conversely, let  $v$  be a child of  $u$ . If it is not maximal in  $M_u$ , then there exists a path  $u \rightsquigarrow v$  different from the arc  $(u, v)$ . Let  $w$  be the parent of  $v$  in this path. Then  $v$  has (at least) two parents,  $u$  and  $w$ , and there is a path  $u \rightsquigarrow w$  in  $N'$ . When we translate this situation to  $N$ , we have essentially four possibilities:



- $v$  is a hybrid node with parents  $u, w$  and there exists a path  $u \rightsquigarrow w$  in  $N$  connecting them.
- $v$  is a hybrid node with parents  $u$  and the tree child  $\bar{w}$  of the hybrid node  $w$ . Then the path  $u \rightsquigarrow w$  in  $N'$  corresponds to a path  $u \rightsquigarrow w$  in  $N$  that, followed by the arc  $(w, \bar{w})$ , yields a path  $u \rightsquigarrow \bar{w}$ .
- $v$  is a hybrid node with parents  $w$  and the tree child  $\bar{u}$  of the hybrid node  $u$ . Then the path  $u \rightsquigarrow w$  in  $N'$  corresponds to a path  $u \rightsquigarrow w$  in  $N$ , and since  $\bar{u}$  is the only child of  $u$ , the latter contains a path  $\bar{u} \rightarrow w$ .
- $v$  is a hybrid node with parents the tree child  $\bar{w}$  of the hybrid node  $w$  and the tree child  $\bar{u}$  of the hybrid node  $u$ . Then the path  $u \rightsquigarrow w$  in  $N'$  corresponds to a path  $u \rightsquigarrow w$  in  $N$ . Arguing as in the last two points (simultaneously), we deduce that there exists a path  $\bar{u} \rightsquigarrow \bar{w}$  in  $N$ .

In all four cases, we obtain a hybrid node of  $N$  and a path connecting two parents of it, which contradicts (4.3.a). This implies that  $v$  must be maximal in  $M_u$ .  $\square$   $\square$

**Theorem 8.** *For every tree-child phylogenetic networks  $N_1$  and  $N_2$  satisfying the weak time consistency property (4.3.a),*

$$N_1 \cong N_2 \text{ if and only if } C(N_1) = C(N_2).$$

*Proof.* Assume that  $C(N_1) = C(N_2)$ , and let  $N'_1 = (V'_1, E'_1)$  and  $N'_2 = (V'_2, E'_2)$  be the contracted versions of  $N_1$  and  $N_2$ . Then  $C(N'_1) = C(N'_2)$ .

Consider the mapping  $f' : V'_1 \rightarrow V'_2$  obtained as the composition

$$V'_1 \xrightarrow{C(N'_1)} C(N'_1) = C(N'_2) \xrightarrow{C(N'_2)^{-1}} V'_2;$$

that is,  $f'$  sends each node  $v \in V'_1$  to the unique node  $f'(v) \in V'_2$  such that  $C^{(N'_1)}(v) = C^{(N'_2)}(f'(v))$ . Since  $C^{(N'_1)}$  and  $C^{(N'_2)}$  are bijective,  $f'$  is bijective. Furthermore,  $v$  is maximal in  $M_u$  if and only if  $f'(v)$  is maximal in  $M_{f'(u)}$  (because these sets are defined through the corresponding clusters, and the path ordering in  $N'_1$  and  $N'_2$  corresponds to the inclusion of clusters). Therefore,  $(u, v) \in E'_1$  if and only if  $(f'(u), f'(v)) \in E'_2$ . So,  $f$  is an isomorphism of DAGs.

Finally,  $u$  is a leaf of a DAG if and only if its cluster is the singleton  $\{u\}$ . This entails that  $f'$  sends leaves to leaves and preserves their labels. Therefore,  $f'$  is an isomorphism of DAGs labeled in  $S$ .

Now, the hybrid nodes in each  $N_i$  are the nodes that have in-degree greater than 1 in the corresponding  $N'_i$ , and  $N_i$  is obtained from  $N'_i$  by adding a single child  $\bar{u}$  to each hybrid node  $u$  and replacing all arcs with tail  $u$  by arcs with tail  $\bar{u}$  (and the same heads). This implies that the mapping

$$f : V_1 \rightarrow V_2$$

that restricts to  $f'$  on  $V'_1$  and that sends each node  $\bar{u}$  in  $V_1 \setminus V'_1$  to the corresponding  $f'(u)$  (that is, to the only child of the image of the parent of  $\bar{u}$ ) is bijective and preserves and reflects the arcs and preserves the leaves' labels. Therefore, it is an isomorphism of phylogenetic networks.  $\square$   $\square$

**Corollary 9.** *The bipartition  $\pi$ , and hence also  $\theta$ ,  $\theta_B$ ,  $\theta_{AB}$ , and  $\Psi$ , satisfy the separation property on the subclass of all tree-child phylogenetic networks where no pair of parents of a hybrid node is connected by a path.*

**Corollary 10.** *For every  $\Upsilon = \pi, \theta, \theta_B, \theta_{AB}, \Psi$ , the mapping*

$$d_{\Upsilon}(N_1, N_2) = \frac{1}{2}(|\Upsilon(T) \setminus \Upsilon(T')| + |\Upsilon(T') \setminus \Upsilon(T)|)$$

*defines a distance on the subclass of all tree-child phylogenetic networks where no pair of parents of a hybrid node is connected by a path.*

## 10 Conclusion

In a series of technical reports and papers culminating in [14], Moret, Nakhleh, Warnow and collaborators have introduced an error metric for phylogenetic networks, with the main goal of comparing reconstructed networks with true ones and to assess in this way the accuracy of phylogenetic network reconstruction algorithms. In this paper we have shown that none of their approaches is free from false equalities: that is, for every one of the metrics they introduce, there turn out to exist pairs of phylogenetic networks in the subclass where the metric is defined, that are non-isomorphic (or, in the case of [14], have non-isomorphic reduced versions) but cannot be distinguished through the metric. The reason for this lack of discriminating power is that non-isomorphic networks in the subclasses under consideration may have the same sets of tripartitions, which are the networks' representations that are compared by this metric. Among these subclasses of networks where the error metric fails, we want to stress the *tree-sibling, strongly time consistent phylogenetic networks* (see Section 7), for which several reconstruction algorithms were recently proposed [9, 10].

We have also shown a subclass of phylogenetic networks, the *tree-child, weakly time consistent phylogenetic networks*, where tripartitions, and even bipartitions in the sense of Bourque-Robinson-Foulds, single out its members, and therefore they can be used to define a true metric. Tree-child phylogenetic networks can be seen as models of reticulate evolution histories where every species other than the extant ones have some descendant through mutation.

Several questions and problems arise as a consequence of our work that are in our current research agenda. On the one hand, what is the discriminating power of tripartitions? Is there a well-defined class of phylogenetic networks where equality of sets, or multisets (sets with repetitions), of tripartitions implies isomorphism? And, what does it really mean, from the topological point of view, to have the same (multi)sets of tripartitions?

On the other hand, it is still necessary to define true distances generalizing the bipartition distance, on more general subclasses of phylogenetic networks than those tree-child, weakly time consistent. We have recently defined one such metric (not based on tripartitions) on the class of *all* tree-child phylogenetic networks [5], but, in the light of [9, 10], we consider a more relevant target the class of all tree-sibling networks.

## References

- [1] Bandelt, H.J.: Phylogenetic networks. Verh. Naturwiss. Ver. Hambg. **34** (1994) 51–71
- [2] Baroni, M., Semple, C., Steel, M.: Hybrids in Real Time. Syst. Biol. **55** (2006) 46–56
- [3] Bourque, M.: Arbres de Steiner et réseaux dont varie l'emplacement de certains sommets. PhD thesis, Univ. de Montréal (1978)
- [4] Burkhardt, F., Smith, S., eds.: The Correspondence of Charles Darwin. Volume 2. Cambridge University Press (1987)

- [5] Cardona, G., Rosselló, F., Valiente, G.: Tree-Child Phylogenetic Networks: Characterization and Distance Measures. Submitted (2007).
- [6] DasGupta, B., He, X., Jiang, T., Li, M., Tromp, J., Wang, L., Zhang, L.: Computing distances between evolutionary trees. In Du, D.Z., Pardalos, P., eds.: Handbook of Combinatorial Optimization. Kluwer Academic Publishers (1998) 35–76
- [7] Doolittle, W.F.: Phylogenetic classification and the universal tree. *Science* **284** (1999) 2124–2128
- [8] Felsenstein, J.: Inferring Phylogenies. Sinauer Associated Inc. (2004)
- [9] Jin, G., Nakhleh, L., Snir, S., Tuller, T.: Maximum likelihood of phylogenetic networks. *Bioinformatics* **22** (2006) 2604–2611
- [10] Jin, G., Nakhleh, L., Snir, S., Tuller, T.: Efficient parsimony-based methods for phylogenetic network reconstruction. *Bioinformatics* **23** (2007) 123–128
- [11] Linder, C.R., Moret, B.M.E., Nakhleh, L., Padolina, A., Sun, J., Tholse, A., Timme, R., Warnow, T.: An error metric for phylogenetic networks. Technical Report TR03-26, University of New Mexico (2003)
- [12] Linder, C.R., Moret, B.M.E., Nakhleh, L., Warnow, T.: Network (reticulate) evolution: Biology, models, and algorithms. Tutorial presented at The Ninth Pacific Symposium on Biocomputing, available online at <http://www.cs.rice.edu/~nakhleh/Papers/psb04.pdf> (2003)
- [13] Moret, B.M.E., Nakhleh, L., Warnow, T.: An error metric for phylogenetic networks. Technical Report TR02-09, University of New Mexico (2002)
- [14] Moret, B.M.E., Nakhleh, L., Warnow, T., Linder, C.R., Tholse, A., Padolina, A., Sun, J., Timme, R.: Phylogenetic networks: Modeling, reconstructibility, and accuracy. *IEEE T. Comput. Biol.* **1** (2004) 13–23
- [15] Nakhleh, L.: Phylogenetic networks. PhD thesis, University of Texas at Austin (2004), available online at <http://bioinfo.cs.rice.edu/Papers/dissertation.pdf>.
- [16] Nakhleh, L., Clement, A., Warnow, T., Linder, C.R., Moret, B.M.E.: Quality measures for phylogenetic networks. Technical Report TR04-06, University of New Mexico (2004)
- [17] Nakhleh, L., Sun, J., Warnow, T., Linder, C.R., Moret, B.M.E., Tholse, A.: Towards the development of computational tools for evaluating phylogenetic network reconstruction methods. In: Proc. 8th Pacific Symp. Biocomputing. (2003) 315–326
- [18] Robinson, D.F., Foulds, L.R.: Comparison of phylogenetic trees. *Mathematical Biosciences* **53** (1981) 131–147
- [19] Semple, C., Steel, M.: Phylogenetics. Volume 24 of Oxford Lecture Series in Mathematics and Its Applications. Oxford University Press (2003)
- [20] Strimmer, K., Moulton, V.: Likelihood analysis of phylogenetic networks using directed graphical models. *Mol. Biol. Evol.* **17** (2000) 875–881
- [21] Strimmer, K., Wiuf, C., Moulton, V.: Recombination analysis using directed graphical models. *Mol. Biol. Evol.* **18** (2001) 97–99

## Appendix

In this Appendix we collect all depictions of phylogenetic networks and their tripartitions.

In graphical representations of phylogenetic networks, and of DAGs in general, hybrid nodes are represented by squares and tree nodes by circles. In those cases where the strong time consistency condition is considered, nodes are labelled with its corresponding  $\tau$  (see Proposition 1) as subscript, to ease the verification of this condition.

In the tables presenting tripartitions we shall make use for simplicity of the following conventions: we only provide the sets  $A$  and  $B$ , as  $C^c$  can be trivially deduced from them; the labels' weights are shown as subscripts, the lack of subscript meaning weight 0; and since the tripartition induced by an arc only depends on its head, we identify the arcs by means of their heads.

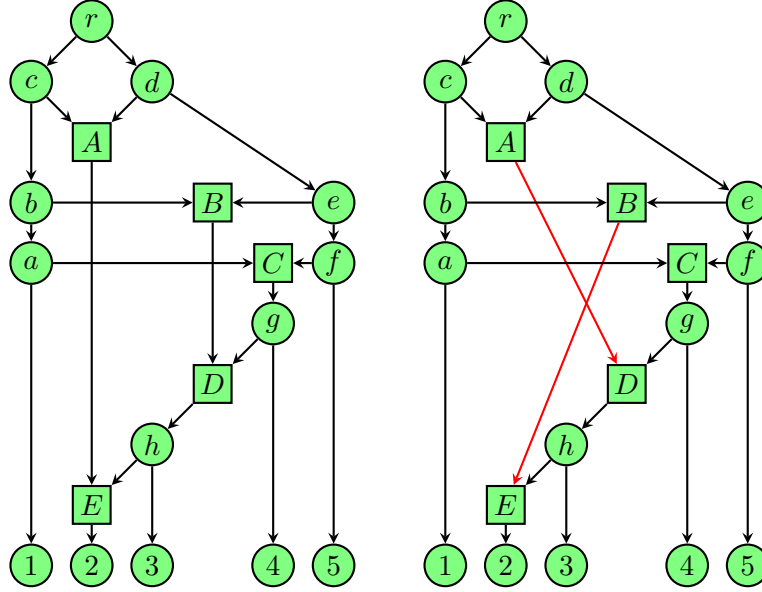


Figure 3: The networks  $N_1$  (left) and  $N_2$  (right)

arc's head	$N_1$		$N_2$	
	$A$	$B$	$A$	$B$
$a$	1	$2_3, 3_2, 4_1$	1	$2_3, 3_2, 4_1$
$b$	1	$2_3, 3_2, 4_1$	1	$2_3, 3_2, 4_1$
$c$	1	$2_3, 3_2, 4_1$	1	$2_3, 3_2, 4_1$
$d$	5	$2_3, 3_2, 4_1$	5	$2_3, 3_2, 4_1$
$e$	5	$2_3, 3_2, 4_1$	5	$2_3, 3_2, 4_1$
$f$	5	$2_3, 3_2, 4_1$	5	$2_3, 3_2, 4_1$
$g$	4	$2_2, 3_1$	4	$2_2, 3_1$
$h$	3	$2_1$	3	$2_1$
$r$	$1, 2_3, 3_2, 4_1, 5$	$\emptyset$	$1, 2_3, 3_2, 4_1, 5$	$\emptyset$
$A$	$\emptyset$	$2_2$	$\emptyset$	$2_3, 3_2$
$B$	$\emptyset$	$2_3, 3_2$	$\emptyset$	$2_2$
$C$	$4_1$	$2_3, 3_2$	$4_1$	$2_3, 3_2$
$D$	$3_1$	$2_2$	$3_1$	$2_2$
$E$	$2_1$	$\emptyset$	$2_1$	$\emptyset$

Table 1: Tripartitions of the networks in Fig. 3

arc's head	$N_3$		$N_4$	
	$A$	$B$	$A$	$B$
$a$	3	$4_3, 8_2, 10_1$	3	$4_3, 8_2, 10_1$
$b$	3	$4_3, 5_3, 6_2, 7_2, 8_2, 9_3, 10_1$	3	$4_3, 5_3, 6_2, 7_2, 8_2, 9_3, 10_1$
$c$	$\emptyset$	$3_1, 4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2$	$\emptyset$	$3_1, 4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2$
$d$	2	$3_1, 4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2$	2	$3_1, 4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2$
$e$	1	$3_1, 4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2$	1	$3_1, 4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2$
$f$	$1, 2, 3_1$	$4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2$	$1, 2, 3_1$	$4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2$
$g$	$11_1, 12, 13$	$4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2$	$11_1, 12, 13$	$4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2$
$h$	13	$4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2, 11_1$	13	$4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2, 11_1$
$i$	12	$4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2, 11_1$	12	$4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2, 11_1$
$j$	$\emptyset$	$4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2, 11_1$	$\emptyset$	$4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2, 11_1$
$k$	11	$4_3, 5_3, 6_2, 7_2, 8_2, 9_3, 10_1$	11	$4_3, 5_3, 6_2, 7_2, 8_2, 9_3, 10_1$
$l$	11	$4_3, 8_2, 10_1$	11	$4_3, 8_2, 10_1$
$m$	$\emptyset$	$4_1, 5_2, 6_1, 7_1, 9_2$	$\emptyset$	$4_2, 5_2, 6_1, 7_1, 8_1, 9_2$
$n$	$\emptyset$	$4_2, 5_2, 6_1, 7_1, 8_1, 9_2$	$\emptyset$	$4_1, 5_2, 6_1, 7_1, 9_2$
$o$	$\emptyset$	$4_1, 5_1$	$\emptyset$	$4_2, 8_1, 9_1$
$p$	$6, 7$	$5_1, 9_1$	$6, 7$	$5_1, 9_1$
$q$	$\emptyset$	$4_2, 8_1, 9_1$	$\emptyset$	$4_1, 5_1$
$s$	6	$5_1$	7	$9_1$
$t$	7	$9_1$	6	$5_1$
$u$	10	$4_2, 8_1$	10	$4_2, 8_1$
$v$	8	$4_1$	8	$4_1$
$r$	$1, 2, 3_1, 4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2, 11_1, 12, 13$	$\emptyset$	$1, 2, 3_1, 4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2, 11_1, 12, 13$	$\emptyset$
$A$	$3_1$	$4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2$	$3_1$	$4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2$
$B$	$11_1$	$4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2$	$11_1$	$4_4, 5_4, 6_3, 7_3, 8_3, 9_4, 10_2$
$C$	$\emptyset$	$4_2, 5_3, 6_2, 7_2, 9_3$	$\emptyset$	$4_3, 5_3, 6_2, 7_2, 8_2, 9_3$
$D$	$\emptyset$	$4_3, 5_3, 6_2, 7_2, 8_2, 9_3$	$\emptyset$	$4_2, 5_3, 6_2, 7_2, 9_3$
$E$	$6_1, 7_1$	$5_2, 9_2$	$6_1, 7_1$	$5_2, 9_2$
$F$	$10_1$	$4_3, 8_2$	$10_1$	$4_3, 8_2$
$G$	$5_1$	$\emptyset$	$9_1$	$\emptyset$
$H$	$9_1$	$\emptyset$	$5_1$	$\emptyset$
$I$	$8_1$	$4_2$	$8_1$	$4_2$
$J$	$4_1$	$\emptyset$	$4_1$	$\emptyset$

Table 2: Tripartitions of the networks in Fig. 4

arc's head	$N_3$	$N_4$
	$RS$	$RS$
$A$	$\{\{1, 3, 4, 5, 6, 7, 8, 9, 10\}, \{3, 4, 5, 6, 7, 8, 9, 10\}\}$	$\{\{1, 3, 4, 5, 6, 7, 8, 9, 10\}, \{3, 4, 5, 6, 7, 8, 9, 10\}\}$
$B$	$\{\{4, 5, 6, 7, 8, 9, 10, 11, 13\}, \{4, 5, 6, 7, 8, 9, 10, 11\}\}$	$\{\{4, 5, 6, 7, 8, 9, 10, 11, 13\}, \{4, 5, 6, 7, 8, 9, 10, 11\}\}$
$C$	$\{\{3, 4, 5, 6, 7, 8, 9, 10\}, \{4, 5, 6, 7, 8, 9, 10, 11\}\}$	$\{\{3, 4, 5, 6, 7, 8, 9, 10\}, \{4, 5, 6, 7, 8, 9, 10, 11\}\}$
$D$	$\{\{3, 4, 5, 6, 7, 8, 9, 10\}, \{4, 5, 6, 7, 8, 9, 10, 11\}\}$	$\{\{3, 4, 5, 6, 7, 8, 9, 10\}, \{4, 5, 6, 7, 8, 9, 10, 11\}\}$
$E$	$\{\{4, 5, 6, 7, 9\}, \{4, 5, 6, 7, 8, 9\}\}$	$\{\{4, 5, 6, 7, 9\}, \{4, 5, 6, 7, 8, 9\}\}$
$F$	$\{\{3, 4, 8, 10\}, \{4, 8, 10, 11\}\}$	$\{\{3, 4, 8, 10\}, \{4, 8, 10, 11\}\}$
$G$	$\{\{4, 5\}, \{5, 6\}\}$	$\{\{4, 8, 9\}, \{7, 9\}\}$
$H$	$\{\{4, 8, 9\}, \{7, 9\}\}$	$\{\{4, 5\}, \{5, 6\}\}$
$I$	$\{\{4, 8, 9\}, \{4, 8, 10\}\}$	$\{\{4, 8, 9\}, \{4, 8, 10\}\}$
$J$	$\{\{4, 5\}, \{4, 8\}\}$	$\{\{4, 5\}, \{4, 8\}\}$

Table 3: Reticulation scenarios of the hybrid nodes of the networks in Fig. 4

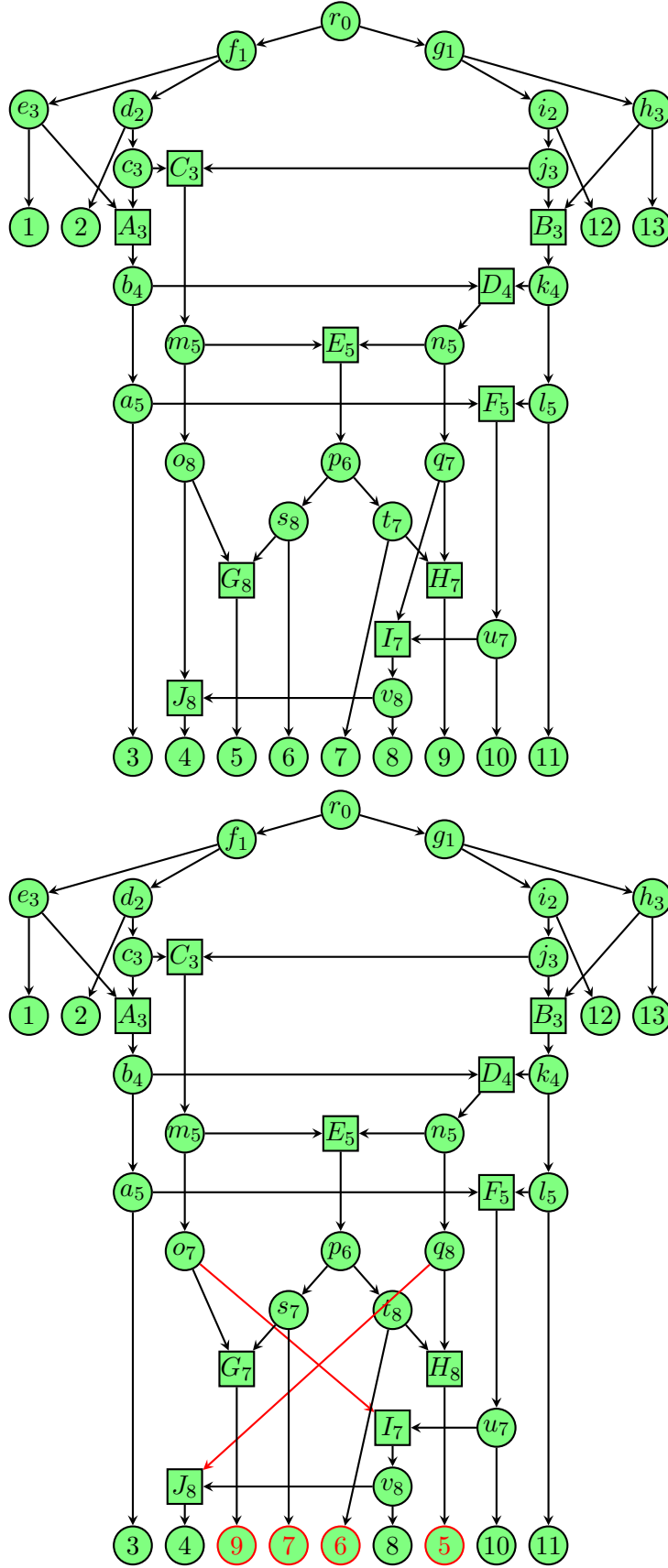


Figure 4: The networks  $N_3$  (up) and  $N_4$  (down)

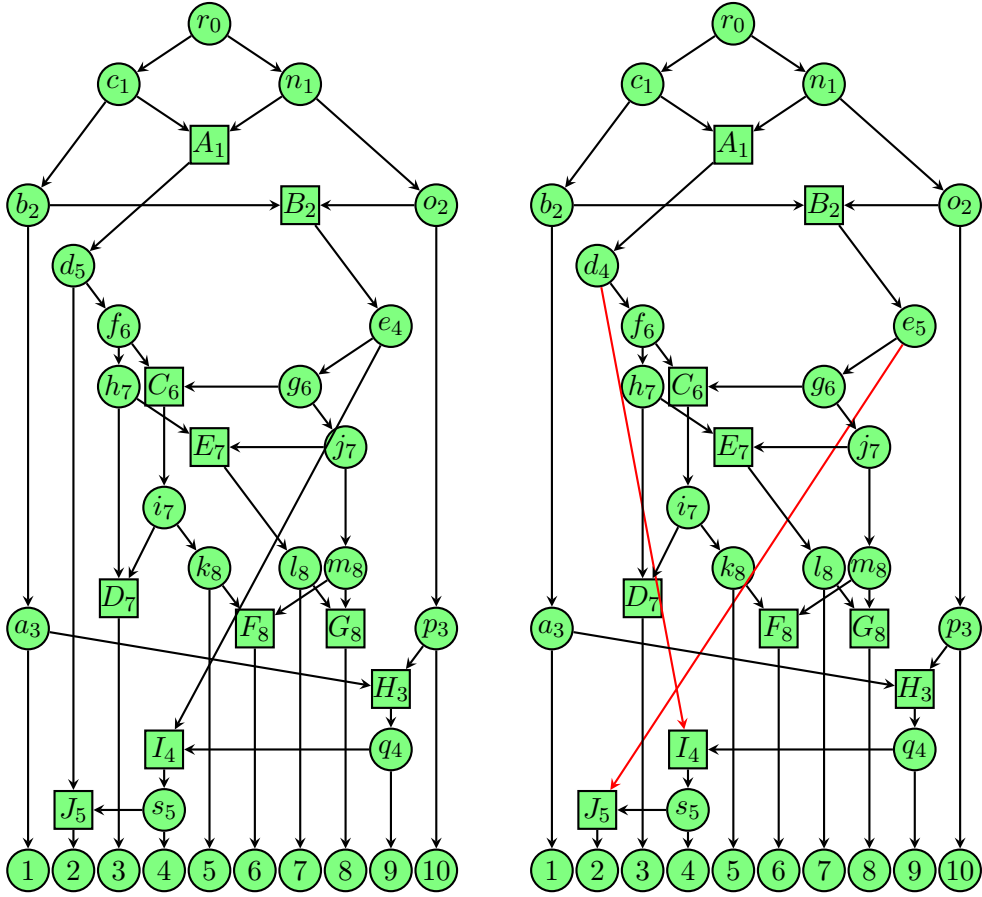


Figure 5: The networks  $N_5$  (left) and  $N_6$  (right)

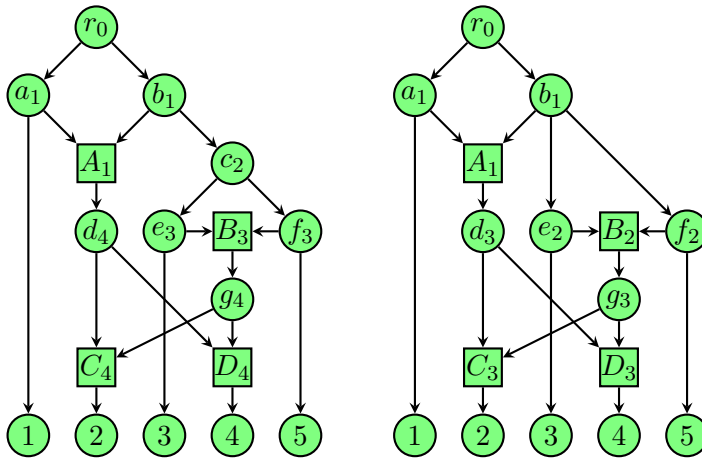


Figure 6: The networks  $N_7$  (left) and  $N_8$  (right)

arc's head	$N_5$		$N_6$	
	$A$	$B$	$A$	$B$
$a$	1	$2_3, 4_2, 9_1$	1	$2_3, 4_2, 9_1$
$b$	1	$2_3, 3_3, 4_2, 5_2, 6_3, 7_2, 8_3, 9_1$	1	$2_3, 3_3, 4_2, 5_2, 6_3, 7_2, 8_3, 9_1$
$c$	1	$2_3, 3_3, 4_2, 5_2, 6_3, 7_2, 8_3, 9_1$	1	$2_3, 3_3, 4_2, 5_2, 6_3, 7_2, 8_3, 9_1$
$d$	$\emptyset$	$2_1, 3_2, 5_1, 6_2, 7_1, 8_2$	$\emptyset$	$2_1, 3_2, 4_1, 5_1, 6_2, 7_1, 8_2$
$e$	$\emptyset$	$2_2, 3_2, 4_1, 5_1, 6_2, 7_1, 8_2$	$\emptyset$	$2_2, 3_2, 5_1, 6_2, 7_1, 8_2$
$f$	$\emptyset$	$3_2, 5_1, 6_2, 7_1, 8_2$	$\emptyset$	$3_2, 5_1, 6_2, 7_1, 8_2$
$g$	$\emptyset$	$3_2, 5_1, 6_2, 7_1, 8_2$	$\emptyset$	$3_2, 5_1, 6_2, 7_1, 8_2$
$h$	$\emptyset$	$3_1, 7_1, 8_2$	$\emptyset$	$3_1, 7_1, 8_2$
$i$	5	$3_1, 6_1$	5	$3_1, 6_1$
$j$	$\emptyset$	$6_1, 7_1, 8_2$	$\emptyset$	$6_1, 7_1, 8_2$
$k$	5	$6_1$	5	$6_1$
$l$	7	$8_1$	7	$8_1$
$m$	$\emptyset$	$6_1, 8_1$	$\emptyset$	$6_1, 8_1$
$n$	10	$2_3, 3_3, 4_2, 5_2, 6_3, 7_2, 8_3, 9_1$	10	$2_3, 3_3, 4_2, 5_2, 6_3, 7_2, 8_3, 9_1$
$o$	10	$2_3, 3_3, 4_2, 5_2, 6_3, 7_2, 8_3, 9_1$	10	$2_3, 3_3, 4_2, 5_2, 6_3, 7_2, 8_3, 9_1$
$p$	10	$2_3, 4_2, 9_1$	10	$2_3, 4_2, 9_1$
$q$	9	$2_2, 4_1$	9	$2_2, 4_1$
$s$	4	$2_1$	4	$2_1$
$r$	$1, 2_3, 3_3, 4_2, 5_2, 6_3, 7_2, 8_3, 9_1, 10$	$\emptyset$	$1, 2_3, 3_3, 4_2, 5_2, 6_3, 7_2, 8_3, 9_1, 10$	$\emptyset$
$A$	$\emptyset$	$2_2, 3_3, 5_2, 6_3, 7_2, 8_3$	$\emptyset$	$2_2, 3_3, 4_2, 5_2, 6_3, 7_2, 8_3$
$B$	$\emptyset$	$2_3, 3_3, 4_2, 5_2, 6_3, 7_2, 8_3$	$\emptyset$	$2_3, 3_3, 5_2, 6_3, 7_2, 8_3$
$C$	$5_1$	$3_2, 6_2$	$5_1$	$3_2, 6_2$
$D$	$3_1$	$\emptyset$	$3_1$	$\emptyset$
$E$	$7_1$	$8_2$	$7_1$	$8_2$
$F$	$6_1$	$\emptyset$	$6_1$	$\emptyset$
$G$	$8_1$	$\emptyset$	$8_1$	$\emptyset$
$H$	$9_1$	$2_3, 4_2$	$9_1$	$2_3, 4_2$
$I$	$4_1$	$2_2$	$4_1$	$2_2$
$J$	$2_1$	$\emptyset$	$2_1$	$\emptyset$

Table 4: Tripartitions of the networks in Fig. 5

arc's head	$N_5$	$N_6$
	$RS$	$RS$
$A$	$\{\{1, 2, \dots, 8, 9\}, \{2, 3, \dots, 9, 10\}\}$	$\{\{1, 2, \dots, 8, 9\}, \{2, 3, \dots, 9, 10\}\}$
$B$	$\{\{1, 2, \dots, 8, 9\}, \{2, 3, \dots, 9, 10\}\}$	$\{\{1, 2, \dots, 8, 9\}, \{2, 3, \dots, 9, 10\}\}$
$C$	$\{\{3, 5, 6, 7, 8\}, \{3, 5, 6, 7, 8\}\}$	$\{\{3, 5, 6, 7, 8\}, \{3, 5, 6, 7, 8\}\}$
$D$	$\{\{3, 7, 8\}, \{3, 5, 6\}\}$	$\{\{3, 7, 8\}, \{3, 5, 6\}\}$
$E$	$\{\{3, 7, 8\}, \{6, 7, 8\}\}$	$\{\{3, 7, 8\}, \{6, 7, 8\}\}$
$F$	$\{\{5, 6\}, \{6, 8\}\}$	$\{\{5, 6\}, \{6, 8\}\}$
$G$	$\{\{7, 8\}, \{6, 8\}\}$	$\{\{7, 8\}, \{6, 8\}\}$
$H$	$\{\{1, 2, 4, 9\}, \{2, 4, 9, 10\}\}$	$\{\{1, 2, 4, 9\}, \{2, 4, 9, 10\}\}$
$I$	$\{\{2, 3, 4, 5, 6, 7, 8\}, \{2, 4, 9\}\}$	$\{\{2, 3, 4, 5, 6, 7, 8\}, \{2, 4, 9\}\}$
$J$	$\{\{2, 4\}, \{2, 3, 5, 6, 7, 8\}\}$	$\{\{2, 4\}, \{2, 3, 5, 6, 7, 8\}\}$

Table 5: Reticulation scenarios of the hybrid nodes of the networks in Fig. 5



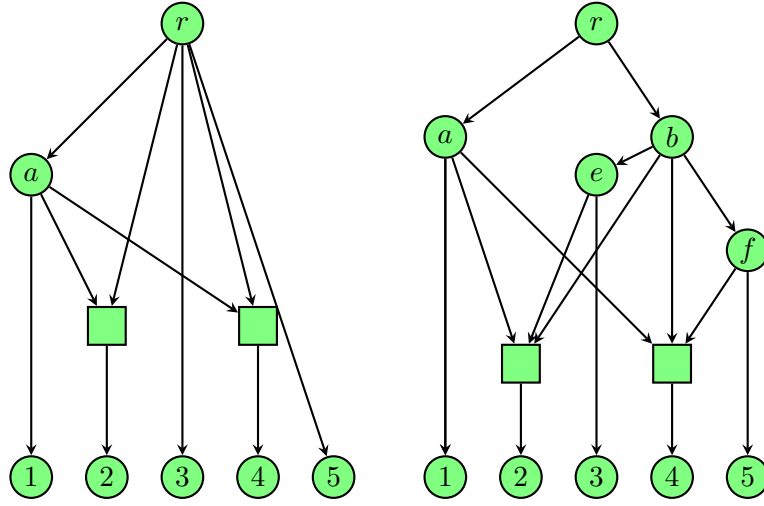


Figure 7: The reduced versions  $R(N_7)$  (left) and  $R(N_8)$  (right) of the networks given in Fig. 6

arc's head	$N_7$		$N_8$	
	$A$	$B$	$A$	$B$
$a$	1	$2_2, 4_2$	1	$2_2, 4_2$
$b$	3, 5	$2_2, 4_2$	3, 5	$2_2, 4_2$
$c$	3, 5	$2_2, 4_2$	—	—
$d$	$\emptyset$	$2_1, 4_1$	$\emptyset$	$2_1, 4_1$
$e$	3	$2_2, 4_2$	3	$2_2, 4_2$
$f$	5	$2_2, 4_2$	5	$2_2, 4_2$
$g$	$\emptyset$	$2_1, 4_1$	$\emptyset$	$2_1, 4_1$
$r$	1, 2, 3, 4, 5	$\emptyset$	1, 2, 3, 4, 5	$\emptyset$
$A$	$\emptyset$	$2_2, 4_2$	$\emptyset$	$2_2, 4_2$
$B$	$\emptyset$	$2_2, 4_2$	$\emptyset$	$2_2, 4_2$
$C$	$2_1$	$\emptyset$	$2_1$	$\emptyset$
$D$	$4_1$	$\emptyset$	$4_1$	$\emptyset$

Table 6: Tripartitions of the networks in Fig. 6

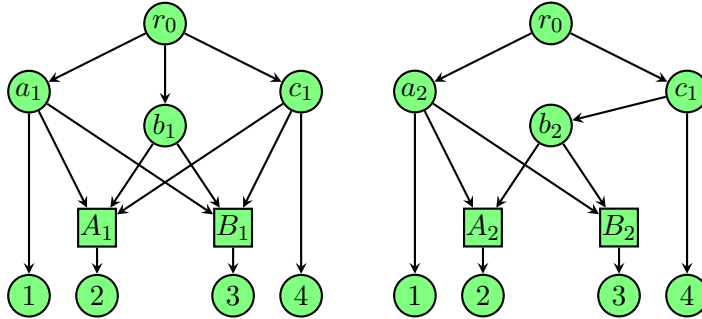


Figure 8: The networks  $N_9$  (left) and  $N_{10}$  (right)

arc's head	$N_9$		$N_{10}$	
	$A$	$B$	$A$	$B$
$a$	1	$2_1, 3_1$	1	$2_1, 3_1$
$b$	$\emptyset$	$2_1, 3_1$	$\emptyset$	$2_1, 3_1$
$c$	4	$2_1, 3_1$	4	$2_1, 3_1$
$r$	$1, 2_1, 3_1, 4$	$\emptyset$	$1, 2_1, 3_1, 4$	$\emptyset$
$A$	$2_1$	$\emptyset$	$2_1$	$\emptyset$
$B$	$3_1$	$\emptyset$	$3_1$	$\emptyset$

Table 7: Tripartitions of the networks in Fig. 8

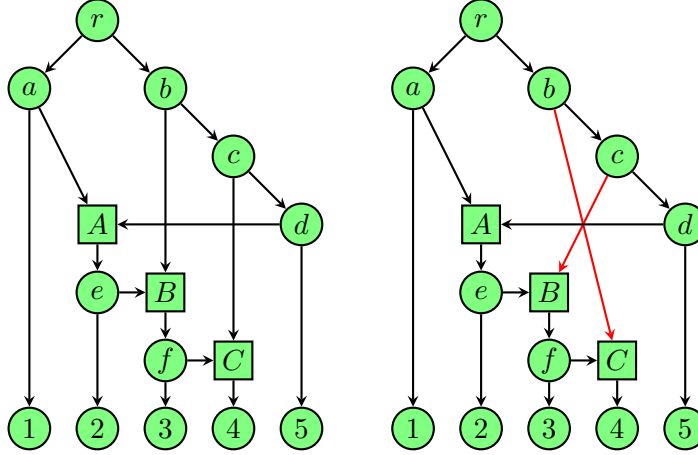


Figure 9: The networks  $N_{11}$  (left) and  $N_{12}$  (right) are tree-child but cannot be distinguished by means of their tripartitions

arc's head	$N_{11}$		$N_{12}$	
	$A$	$B$	$A$	$B$
$a$	1	$2_1, 3_2, 4_3$	1	$2_1, 3_2, 4_3$
$b$	5	$2_1, 3_2, 4_3$	5	$2_1, 3_2, 4_3$
$c$	5	$2_1, 3_2, 4_3$	5	$2_1, 3_2, 4_3$
$d$	5	$2_1, 3_2, 4_3$	5	$2_1, 3_2, 4_3$
$e$	2	$3_1, 4_2$	2	$3_1, 4_2$
$f$	3	$4_1$	3	$4_1$
$r$	$1, 2_1, 3_2, 4_3, 5$	$\emptyset$	$1, 2_1, 3_2, 4_3, 5$	$\emptyset$
$A$	$2_1$	$3_2, 4_3$	$2_1$	$3_2, 4_3$
$B$	$3_1$	$4_2$	$3_1$	$4_2$
$C$	$4_1$	$\emptyset$	$4_1$	$\emptyset$
$RS$			$RS$	
$A$	$\{\{1, 2, 3, 4\}, \{2, 3, 4, 5\}\}$		$\{\{1, 2, 3, 4\}, \{2, 3, 4, 5\}\}$	
$B$	$\{\{2, 3, 4\}, \{2, 3, 4, 5\}\}$		$\{\{2, 3, 4\}, \{2, 3, 4, 5\}\}$	
$C$	$\{\{3, 4\}, \{2, 3, 4, 5\}\}$		$\{\{3, 4\}, \{2, 3, 4, 5\}\}$	

Table 8: Tripartitions and reticulation scenarios of the tree-child phylogenetic networks in Fig. 9